Strong Consistency in a Geo-Replicated File System

Benjamin Bengfort
University of Maryland
bengfort@cs.umd.edu

Pete Keleher
University of Maryland
keleher@cs.umd.edu

Implementation

The FluidFS File System is implemented in Golang as a filesystem in user space using the `bazil.org/fuse` implementation in pure Go (without libfuse bindings). This dependency is a custom implementation of the kernel-userspace communication protocol inspired by the FUSE library.

FluidFS mounts one or more mount points defined in a host-specific, `fstab`-like configuration file. Each mount point constructs an independent file system from the perspective of the kernel, whose top level is the root directory where the FS is mounted. FluidFS, however treats each mount point as a subdirectory of an abstract global file system, specified by a unique *prefix* or TLD (top level directory) that must also be identified in the configuration file. When two mount points are specified with the same prefix (either locally or on different hosts), FluidFS treats each mount point as a partial replica of the global file system.

FluidFS is a virtual distributed file system that does not manage a disk directly but instead primarily manages objects (files). A file is a collection of version metadata and binary blobs that are stored separately. A *file version* is represented by a single piece of meta data that contains two Lamport scalar compound numbers: the version number and the parent of the current version.

References